Niveau(x) d'enseignement: 1 Numéro EDUQ : 8670180
Centre d'intérêt: 640

L'Utilisation dans un contexte d'évaluation formative d'un test généré par ordinateur à partir d'une banque d'items / par Denis Rhéaume... [et al.]. -- Sainte-Foy : INRS-Education, 1983.
6 f. ; 28 cm.

Code : C-110.

Vedette(s)-secondaire(s): Rhéaume, Denis, 1954- * INRS-Education

---

Descripteur(s): Application pédagogique de l'ordinateur * Elaboration de tests * Evaluation formative * Méthode d'évaluation * Primaire (Ordre d'enseignement) * Secondaire (Ordre d'enseignement)

Dans le cadre de travaux relatifs aux applications pédagogiques du micro-ordinateur, des chercheurs de l'INRS-Education se sont donné comme objectif de développer un système informatique permettant la création d'un test à partir d'une banque d'items et sa passation au terminal. Le but de cet article est, d'une part, de faire connaître les principales caractéristiques du système développé et de justifier certaines décisions qui ont été prises au moment de sa conception et, d'autre part, de montrer comment un tel système peut parfois s'avérer un outil précieux dans un contexte d'évaluation formative. De plus, certains éléments de la démarche prévue pour l'évaluation du système sont présentés.

Type de document: r

Pg 73

EDUQ 8670180

Denis Rhéaume, Normand Dufour, Marie Makdissi et Yves Bégin.

UTILISATION DANS UN CONTEXTE D'EVALUATION FORMATIVE D'UN TEST GENERE PAR ORDINATEUR A PARTIR D'UNE BANQUE D'ITEMS.

INRS-Education, C-110, 1983

EDUQ 8670180

L'UTILISATION DANS UN CONTEXTE D'ÉVALUATION FORMATIVE D'UN TEST GÉNÉRÉ PAR ORDINATEUR À PARTIR D'UNE BANQUE D'ITEMS

par

Denis Rhéaume
Normand Dufour
Marie Makdissi
Yves Bégin

C-110

INRS-Éducation
2383 chemin Ste-Foy
Sainte-Foy
G1V 1T1

Novembre 1983

L'UTILISATION DANS UN CONTEXTE D'ÉVALUATION FORMATIVE D'UN TEST GÉNÉRÉ PAR ORDINATEUR À PARTIR D'UNE BANQUE D'ITEMS

D. Rhéaume, N. Dufour, M. Makdissi, Y. Bégin

Institut national de la recherche scientifique (Éducation),
Québec G1V 1T1

Dans le cadre de travaux relatifs aux applications pédagogiques de l'ordinateur personnel, des chercheurs de l'INRS-Éducation se sont donnés comme objectif de développer un système informatique permettant la génération d'un test à partir d'une banque d'items et sa passation au terminal. En plus de la conception et de la réalisation proprement dite du système, une phase d'évaluation a été prévue laquelle se déroulera durant l'année scolaire 1983-1984.

Le but de cette communication est d'une part de faire connaître les principales caractéristiques du système développé et de justifier certaines décisions qui ont été prises au moment de sa conception et d'autre part, de montrer comment un tel système peut parfois s'avérer un outil précieux dans un contexte d'évaluation formative. De plus, certains éléments de la démarche prévue pour l'évaluation du système seront présentés.

Avant d'aller plus loin, il est utile de préciser que les travaux auxquels réfère cette communication s'étendent sur une période de deux ans. La première année a été consacrée à la production du système. Cette année, le système pourra être utilisé dans des classes à titre expérimental et l'équipe de recherche met actuellement au point un processus d'évaluation dont les résultats seront colligés dans un rapport d'ici un an.

Le système développé permet d'utiliser l'ordinateur personnel pour:

(1) construire une banque d'items;

(2) générer et imprimer un test en choisissant au hasard des items de la banque;

(3) présenter à l'écran un test composé d'items choisis au hasard permettant ainsi à l'étudiant[1] d'utiliser le terminal pour répondre.

Ce système peut servir aussi bien à l'école primaire ou secondaire qu'au cégep ou à l'université. Il n'est limité ni à une matière en particulier ni à un groupe de matières et son utilisation est possible à chaque fois que les items choisis pour la mesure du rendement sont des items à choix de réponses ou des items qui appellent une réponse très courte et une correction objective.

Qui peut être l'utilisateur d'un tel système? L'équipe de recherche a voulu produire un système suffisamment simple pour en permettre l'utilisation par des individus non initiés à cette technologie nouvelle que représente l'ordinateur personnel. Dans les faits, les membres de l'équipe ont accordé beaucoup d'attention à la qualité de la relation entre l'utilisateur et l'ordinateur.

Concrètement, les utilisateurs de ce système pourraient très bien être l'enseignant qui désire s'en servir avec ses propres étudiants comme un groupe d'enseignants désirant mettre en commun leurs efforts pour la production d'une banque d'items. Enfin, il pourrait être avantageusement exploité par des individus désirant produire des tests mis par la suite à la disposition des personnes intéressées. Le système développé n'est rien de plus mais rien de moins qu'un logiciel pour la production de tests.

QUATRE CARACTÉRISTIQUES IMPORTANTES

Déjà, il existe plusieurs systèmes poursuivant des objectifs semblables aux nôtres et nous en étions conscients au point de départ. Nous avons néanmoins voulu doter le logiciel développé à l'INRS-Éducation de certaines caractéristiques que nous jugions importantes mais que nous ne retrouvions pas dans la plupart des autres systèmes.

Cet exposé met en évidence quatre de ces caractéristiques.

---

1 Dans ce texte, le mot étudiant réfère aussi bien aux étudiantes qu'aux étudiants et le mot enseignant réfère aussi bien aux enseignantes qu'aux enseignants.

1. Le système permet un échantillon stratifié d'items

En fait, à chaque fois que l'utilisateur introduit un item dans la banque, il doit lui attribuer un numéro de catégorie. Une catégorie représente un groupe d'items, un sous-ensemble de la banque; chaque catégorie peut correspondre à une section de la matière, à un objectif pédagogique ou à tout autre classement jugé utile.

Lors de la génération d'un test, l'utilisateur peut déterminer le nombre désiré d'items provenant de chaque catégorie, ce nombre pouvant varier d'une catégorie à l'autre; le test devient alors la juxtaposition de plusieurs sous-tests composés d'items choisis au hasard et c'est ce qui nous amène à parler d'échantillon stratifié d'items. De plus, le système permet aussi à l'utilisateur de ne pas tenir compte du classement des items en catégories et d'obtenir un seul échantillon d'items.

La présence de l'échantillon stratifié comporte plus d'un avantage. Premièrement, il n'est plus nécessaire d'assurer une relation parfaite entre l'importance d'une catégorie et le nombre d'items disponibles dans la banque pour cette catégorie. Deuxièmement, l'utilisateur peut faire varier l'importance rattachée à une catégorie selon les circonstances et cela sans devoir ajouter ou retrancher des items dans la banque. Troisièmement, cet échantillon stratifié assure une meilleure représentativité de l'ensemble des items formant le test.

Le système d'échantillon stratifié est relativement souple. Le nombre de catégories différentes peut atteindre vingt-cinq; le nombre d'items pouvant appartenir à une catégorie n'est pas limité, sinon par la taille de la banque elle-même et l'utilisateur peut changer en tout temps la catégorie d'appartenance d'un item.

2. Le système permet de considérer certaines formes d'items qui appellent une réponse courte et d'aller ainsi au-delà des items à choix de réponses

Les membres de l'équipe étaient conscients au point de départ qu'il était beaucoup plus facile de produire un système ne permettant de considérer que des items pour lesquels l'étudiant choisit la lettre ou le chiffre correspondant à sa réponse. Nous avons cependant voulu bénéficier de la possibilité qu'offre l'ordinateur pour la correction des items qui appellent une réponse courte et non uniquement un choix de réponses.

Le système prévoit donc sept types d'items:

type 1: choix multiples

type 2: vrai-faux

type 3: réponse simple (forme 1)

L'étudiant répond en donnant une chaîne de caractères (exemple: un mot, un chiffre). Ici une chaîne désigne toute suite de lettres ou de chiffres juxtaposés. Aucune réponse n'est donc proposée. Une seule chaîne est considérée étant la bonne réponse.

type 4: réponse simple (forme 2)

Encore une fois, l'étudiant répond en donnant une seule chaîne de caractères. Cette fois, deux ou trois chaînes peuvent être considérées comme étant la bonne réponse, ce qui permet une correction plus subtile. Une telle possibilité est parfois essentielle. Voici un exemple: si l'on demande à l'étudiant de traduire le mot "automne" en anglais, deux réponses doivent être considérées comme acceptables, "fall" et "autumn".

type 5: réponse plus longue (forme 1)

Cette fois, la réponse de l'étudiant peut comporter plusieurs chaînes de caractères mais elle ne doit pas dépasser 40 caractères. Il s'agit donc toujours d'items qui appellent une réponse courte. Aucune réponse n'est proposée. Pour être jugée bonne, la réponse doit contenir une chaîne donnée laquelle est fournie à l'ordinateur comme clé de correction. Voici un exemple d'utilisation: si l'on demande de nommer la rivière la plus importante qui serpente le territoire de la ville de Québec, l'on pourra considérer qu'il y a bonne réponse dès que l'ordinateur rencontrera la chaîne "charles" laquelle doit nécessairement être contenue dans la réponse de l'étudiant qui sait que cette rivière est la rivière Saint-Charles.

type 6: réponse plus longue (forme 2)

Encore une fois, la réponse de l'étudiant peut comporter plusieurs chaînes de caractères mais ne doit pas dépasser 40 caractères.

Cette fois cependant, l'ordinateur peut prendre en considération deux ou trois chaînes comme clé de correction; la réponse doit contenir l'une ou l'autre des deux ou trois chaînes pour être considérée comme bonne.

type 7: estimé numérique

Pour ce septième et dernier type, encore une fois, aucune réponse n'est proposée à l'étudiant. Sa réponse doit être un chiffre; ce chiffre peut être accompagné ou non d'un signe; il peut comporter une partie décimale ou non. Pour ce type, l'ordinateur considère qu'il y a bonne réponse si la réponse fournie se situe entre une limite inférieure et une limite supérieure lesquelles sont fournies par l'utilisateur comme clé de correction.

C'est surtout durant la phase d'évaluation que seront identifiés les problèmes et les avantages que soulève l'utilisation des divers types d'items. Par ailleurs, le type de l'item n'a de l'importance que lorsque l'on prévoit que le test sera présenté de façon interactive au terminal. En effet, si le test est imprimé, le problème de la correction par ordinateur ne se pose plus, à moins qu'on ne tienne à utiliser le lecteur optique. Pour le test imprimé, la prise en compte par le système d'une variété de types d'items est nécessairement avantageuse.

3. Le système comporte un dispositif faisant en sorte que des items jugés incompatibles ne fassent pas partie d'un même test

Notre expérience ayant trait à l'utilisation de banques d'items pour la génération de tests nous a graduellement amenés à voir que, dans plusieurs cas, la coapparition de certains items dans un même test engendrait certains problèmes de validité. Mentionnons les deux problèmes les plus courants qui peuvent survenir. Premièrement, la présence d'un item donné dans le test peut contenir un indice important de la réponse à un autre item, ce qui est évidemment habituellement non désirable. Deuxièmement, il peut arriver que l'on désire placer dans une même banque deux items de forme différente mais dont la similitude de contenu rend injustifiable leur coapparition.

Les problèmes évoqués plus haut peuvent facilement être résolus en décidant de ne jamais mettre dans une même banque des items jugés incompatibles et d'exclure ainsi systématiquement certains items pour cette raison. Cependant, nous avons préféré solutionner ces problèmes en incluant dans notre système un dispositif permettant que des items de la banque jugés incompatibles ne fassent pas partie d'un même test. Cette façon de faire devrait permettre de conserver dans les banques d'items plusieurs items de qualité qui auraient normalement dû être éliminés.

L'introduction d'un tel dispositif, simple en apparence, a engendré au niveau de la programmation, plusieurs difficultés qui ont par la suite été surmontées.

Nous verrons mieux pendant la période d'évaluation l'utilisation qui peut être faite d'un tel dispositif. Mentionnons enfin que ce dispositif est utile non seulement lorsque le test est imprimé mais aussi lorsqu'il est présenté à l'écran.

4. Le système permet à l'étudiant de revenir à un item antérieur pour répondre ou changer sa réponse comme s'il s'agissait d'un test écrit

Cette caractéristique du système ne concerne que le cas où le test est présenté à l'écran. Au point de départ, nous avions le choix entre deux avenues. La première, la plus utilisée par les sytèmes semblables aux nôtres, consistait à fournir à l'étudiant un feed-back immédiat lorsque sa réponse est donnée. La deuxième, celle que nous avons retenue, consiste à permettre de répondre à tous les items du test avant de donner le résultat. Nous avons choisi cette deuxième avenue parce que c'est celle qui de toute évidence s'apparente le plus au test écrit; de plus nous pensons que le feed-back à la fin du test permet d'obtenir un résultat peut-être plus valide dans certains cas, étant donné que le feed-back après chaque item équivaut à fournir de l'information durant la passation du test.

Cette option étant prise, il devenait souhaitable d'exiger que l'ordinateur offre au minimum tous les avantages du test écrit et, d'une façon particulière, la possibilité pour l'étudiant de revenir à un item antérieur pour répondre ou changer sa réponse.

Avec notre système, l'étudiant peut identifier ses bonnes réponses et ses erreurs après avoir connu son résultat.

TROIS CARACTÉRISTIQUES SECONDAIRES

Il importe maintenant de mentionner trois caractéristiques secondaires du système en développement. Elles sont qualifiées de secondaires uniquement parce qu'elles ne concernent

pas proprement le domaine touché par notre application, soit la mesure du rendement scolaire. Elles sont toutefois bien importantes dans notre esprit.

Premièrement, le système développé permet d'utiliser pour le texte des items les minuscules et les accents sur les minuscules. Il respecte donc en bonne partie les exigences de la langue française, même si les accents sur les majuscules ne sont pas encore possibles.

Deuxièmement, le texte des items introduits dans la banque peut être sensiblement plus long que ce qu'offrent plusieurs systèmes comparables au nôtre. Il peut atteindre plus de 400 caractères. Cette possibilité permet de rendre utilisables bon nombre d'items qui devraient être rejetés avec d'autres systèmes; elle est particulièrement utile pour les items à choix multiples lorsque les réponses proposées à l'étudiant sont longues.

Troisièmement, le système développé comporte un dispositif simple d'utilisation mais relativement puissant permettant de modifier ou de corriger à n'importe quel moment le texte d'un item. Déjà, nous avons l'évidence de la rentabilité d'un tel dispositif, et ce, particulièrement lorsque les items sont longs.

LIMITES DU SYSTÈME

Le système développé n'est pas sans limites, loin de là. La plupart d'entre elles ont surtout trait à l'équipement utilisé. L'utilisation du logiciel développé exige, un ordinateur personnel Apple II de 48K, un lecteur de disquette conventionnel, un imprimante et un écran de 40 colonnes.

Actuellement, la capacité maximale de la banque est de 160 items. Le temps de réponse de l'ordinateur est parfois lent pour certaines opérations. L'utilisation de 40 colonnes pour l'impression du test pose certains problèmes de gaspillage et même d'élégance. La taille d'un test ne doit pas dépasser 50 items, et ce, peu importe leur longueur.

Il faut également dire que l'entrée dans la banque d'items dont la forme graphique est complexe pose actuellement des problèmes insurmontables; même l'utilisation d'indices et d'exposants cause des difficultés très considérables. Dans un certaine mesure, l'utilisateur peut contourner ce problème en utilisant un support visuel non informatisé.

D'une certaine façon, nous considérons que nos principales limites ont trait à l'équipement mais que celles-ci seront rapidement surmontées grâce aux développements technologiques déjà envisagés. Il est par ailleurs intéressant de voir tout ce qui peut déjà être fait avec un équipement de potentiel réduit.

UTILITÉ DU SYSTÈME

Quelques mots maintenant concernant l'utilité du système développé. Disons d'abord que c'est surtout durant la phase d'évaluation du projet que nous serons le mieux en mesure de voir quelles sont les possibilités du système qui apparaissent les plus utiles.

Avant tout, notre système en est un qui permet de construire une banque d'items et de construire un test en choisissant au hasard des items de cette banque. Pour ce faire, l'ordinateur n'est pas requis; il permet cependant de faire en très peu de temps ce qui serait normalement très long et à peine recommandable. L'utilisation d'une banque d'items et d'un programme de génération de tests rend possible la production presque automatique de formes parallèles d'un test. Un étudiant peut avoir la possibilité de passer plus d'une fois un test référant au même contenu mais comportant des items différents.

Là où l'ordinateur est essentiel, c'est pour la présentation du test à l'écran et la correction automatisée. Avec un tel outil, si le contexte de la classe le permet, cela signifie qu'un étudiant peut se présenter à l'ordinateur au moment qu'il choisit et subir un test répondant à certaines caractéristiques précises. La disponibilité d'un tel outil rend davantage possible, nous aimerions croire, le développement d'habitudes d'auto-évaluation. L'étudiant peut se présenter à plusieurs reprises pour un même test. Il est le seul à connaître son résultat et à voir ses erreurs, ce qui peut dans certains cas être psychologiquement très confortable. Disons que même s'il ne permet pas le diagnostic, le système actuellement en développement pourrait s'avérer un outil approprié et apprécié dans un contexte d'évaluation formative et que c'est en partie ce que nous entendons vérifier lors de la deuxième partie de nos travaux.

Les composantes du système qui concernent la gestion de la banque et l'impression de tests conviennent aussi bien dans un contexte d'évaluation formative que dans un contexte d'évaluation sommative. Il est cependant important de dire que ce n'est pas le cas de la partie qui concerne la passation de tests au terminal. Elle a en effet été conçue d'abord et avant tout pour l'évaluation formative. Nous ne serions pas portés à recommander l'utilisation du système pour la passation de tests au terminal dans un contexte d'évaluation sommative où il s'agit d'attribuer à un étudiant

une note qui aboutira à son dossier. Plusieurs problèmes pourraient alors se poser. D'abord, le système ne permet pas dans sa forme actuelle de conserver les résultats en mémoire. Les chances pour que plusieurs items reviennent d'une occasion à l'autre et d'un étudiant à l'autre risquent d'être un peu trop élevées pour qu'il demeure justifié de parler de mesures quasi-parallèles. Aussi, si les items de la banque sont utilisés dans un contexte d'évaluation formative et dans un contexte d'évaluation sommative, il y a le danger qu'un même étudiant en vienne à trop bien connaître le contenu de la banque. Également, il faut mentionner que si un test doit être présenté de façon inter-active à un grand nombre d'étudiants et ce, durant une courte période de temps, certains problèmes d'organisation liés à la disponibilité d'équipement peuvent survenir. Dans un autre ordre d'idée, il faut ajouter que le nombre d'items faisant partie d'un test devrait aussi être relativement élevé pour que les possibilités d'obtenir des résultats fiables soient suffisamment élevées. Ce problème d'échantillonnage, typique de l'utilisation de banques d'items, mériterait un examen attentif qu'il n'est pas à propos de faire ici.

Le système développé convient beaucoup plus à un contexte d'évaluation formative, du moins, si l'on se réfère à la possibilité de la passation du test au terminal. L'évaluation formative visant une vérification périodique de la qualité de l'apprentissage au cours du processus d'apprentissage, le système peut offrir à l'étudiant ou à l'étudiante une excellente occasion de recevoir rapidement un feed-back, et ce, au moment voulu. La possibilité de s'impliquer aussi simplement dans un processus d'évaluation formative peut revêtir des aspects psychologiques ou pédagogiques intéressants et c'est notamment ce que nous entendons voir durant la dernière partie de nos travaux.

LA DÉMARCHE D'ÉVALUATION

Le moment est maintenant venu de présenter quelques éléments du processus prévu pour l'évaluation du système. Il est d'abord important de mentionner qu'avant même de commencer les travaux, il apparaissait comme essentiel pour les membres de l'équipe de procéder non seulement à la production d'un système informatique mais également à son évaluation.

Deux aspects déterminants du contexte de l'évaluation envisagée méritent d'être mentionnés. Premièrement, les ressources humaines et matérielles dont nous pouvions disposer ne permettaient qu'une évaluation à une échelle très restreinte, c'est-à-dire notamment dans un très petit nombre de classes. Deuxièmement, le fait que nous soyons à la fois les concepteurs et les évaluateurs du système nous amenait à ne pas retenir certains buts possibles d'une évaluation de logiciels et à profiter de cette situation pour privilégier d'autres buts.

Nous en sommes venus à envisager une démarche d'évaluation ayant comme objectif principal d'identifier des faiblesses et des améliorations possibles, dans le but d'améliorer le système et la présentation que nous en faisons. Notre point de vue n'est donc pas de statuer au bout de la course sur la valeur du système développé; d'une certaine façon, nous l'avons développé parce que nous y croyions et nous sommes enclins à continuer d'y croire.

Il nous a semblé toutefois que nous étions bien placés pour mettre en place une démarche d'évaluation qui aurait comme objectif d'identifier des faiblesses et des améliorations à apporter. Une telle étude pouvait être faite sur une échelle plutôt restreinte et nous avions les ressources nécessaires pour la réaliser. Par ailleurs, le fait que nous soyons à la fois concepteurs et évaluateurs ne pose aucun problème pour ce type d'étude; au contraire, notre connaissance du système facilite le déroulement de l'évaluation et l'apport des correctifs jugés requis.

Comme cela a déjà été mentionné, nous visons à améliorer non seulement le système mais aussi la présentation écrite qui en est faite. Ce deuxième volet prend à nos yeux une importance considérable dans la mesure où le système informatique proprement dit et la présentation qui en est faite forment un tandem indissociable dans la prise en compte de l'autonomie de l'utilisateur éventuel.

La démarche d'évaluation aura donc l'allure suivante. Le système sera présenté à un nombre restreint d'utilisateurs potentiels. Certaines informations seront recueillies dans un contexte d'utilisation du système au laboratoire, d'autres proviendront de l'utilisation libre et autonome du système par l'enseignant participant au projet, et enfin, d'autres observations proviendront de l'utilisation du système dans le contexte réel de la classe; ce troisième type d'informations concerne évidemment uniquement la composante du système qui a trait à la passation d'un test au terminal.

Puisqu'il s'agit d'évaluation, il va un peu de soi qu'il soit également question de critères d'évaluation. À ce sujet, il est d'ores et déjà permis de prévoir que l'attention sera centrée surtout sur la quantité et la nature des ennuis techniques dus à la programmation, la qualité et l'efficacité de la présentation écrite du système, la qualité de la communication entre l'usager et l'ordinateur et l'efficacité du système.

En plus de recueillir des informations ayant trait à l'évaluation proprement dite du système, nous profitons du fait qu'il soit utilisé de façon expérimentale par quelques enseignants et dans quelques classes pour noter certaines observations relatives à l'appréciation du système par les utilisateurs, à l'utilité perçue pour les diverses fonctions spécifiques intégrées au système et, plus globalement, à l'appropriation du système. Mentionnons enfin que les modalités concernant la cueillette des informations désirées ne sont pas pour le moment toutes établies.

Nous espérons que cette démarche d'évaluation permettra de constater que le système et sa présentation écrite ne posent pas de problèmes sérieux même si elle peut entraîner des modifications mineures susceptibles de conduire à un produit final réellement diffusable.